

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of:

MI-SUK LEE, ET AL.

Application No.:

Filed:

For: **Voice Activity Detector and Voice
Activity Detection Method Using
Complex Laplacian Model**

Art Group:

Examiner:

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

REQUEST FOR PRIORITY

Sir:

Applicant respectfully requests a convention priority for the above-captioned application, namely:

COUNTRY	APPLICATION NUMBER	DATE OF FILING
Korea	10-2002-0083728	24 December 2002

☒ A certified copy of the document is being submitted herewith.

Respectfully submitted,

Blakely, Sokoloff, Taylor & Zafman LLP

Dated: _____

10/30/03

Eric S. Hyman, Reg. No. 30,139

12400 Wilshire Blvd., 7th Floor
Los Angeles, California 90025
Telephone: (310) 207-3800

대한민국 특허청

KOREAN INTELLECTUAL PROPERTY OFFICE

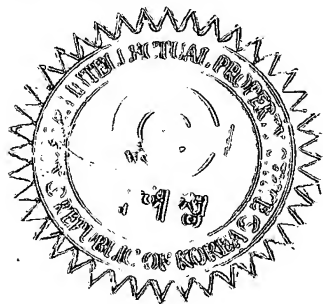
별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto
is a true copy from the records of the Korean Intellectual
Property Office.

출원번호 : 10-2002-0083728
Application Number

출원년월일 : 2002년 12월 24일
Date of Application DEC 24, 2002

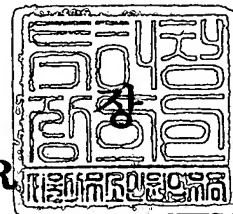
출원인 : 한국전자통신연구원
Applicant(s) Electronics and Telecommunications Research Institute



2003 07 월 28 일

특 허 청

COMMISSIONER



【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	0023
【제출일자】	2002.12.24
【발명의 명칭】	복소수 라플라시안 통계모델을 이용한 음성 검출기 및 음성 검출 방법
【발명의 영문명칭】	A Voice Activity Detector Employing Complex Laplacian Model
【출원인】	
【명칭】	한국전자통신연구원
【출원인코드】	3-1998-007763-8
【대리인】	
【명칭】	유미특허법인
【대리인코드】	9-2001-100003-6
【지정된변리사】	이원일
【포괄위임등록번호】	2001-038431-4
【발명자】	
【성명의 국문표기】	이미숙
【성명의 영문표기】	LEE,MI SUK
【주민등록번호】	700310-2451057
【우편번호】	302-739
【주소】	대전광역시 서구 만년동 상아아파트 109동 1107호
【국적】	KR
【발명자】	
【성명의 국문표기】	황대환
【성명의 영문표기】	HWANG,DAE HWAN
【주민등록번호】	630716-1010012
【우편번호】	305-729
【주소】	대전광역시 유성구 전민동 나래아파트 103동 905호
【국적】	KR
【발명자】	
【성명의 국문표기】	장준혁
【성명의 영문표기】	CHANG,JOON HYUK

【주민등록번호】	730916-1675819
【우편번호】	137-070
【주소】	서울특별시 서초구 서초동 1665-22번지 B01호
【국적】	KR
【발명자】	
【성명의 국문표기】	김남수
【성명의 영문표기】	KIM,NAM S00
【주민등록번호】	651018-1046416
【우편번호】	137-752
【주소】	서울특별시 서초구 방배3동 경남아파트 8동 103호
【국적】	KR
【심사청구】	청구
【취지】	특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인 유미특허법인 (인)
【수수료】	
【기본출원료】	19 면 29,000 원
【가산출원료】	0 면 0 원
【우선권주장료】	0 건 0 원
【심사청구료】	5 항 269,000 원
【합계】	298,000 원
【감면사유】	정부출연연구기관
【감면후 수수료】	149,000 원
【기술이전】	
【기술양도】	희망
【실시권 허여】	희망
【기술지도】	희망
【첨부서류】	1. 요약서·명세서(도면)_1통

【요약서】**【요약】**

본 발명은 복소수 라플라시안 통계 모듈을 이용한 음성 검출기에 관한 것이다. 본 발명의 음성검출기는 입력된 음성에 대하여 고속 푸리에 변환을 수행하여 시간 영역의 음성신호를 주파수 영역에서 분석할 수 있도록 하는 고속 주파수 푸리에 변환기, 상기 고속 주파수 푸리에 변환기로부터 출력되는 주파수 영역 상의 잡음에 오염된 음성으로부터 잡음 신호의 파워를 추정하는 잡음 파워 추정부, 상기 잡음 파워 추정부로부터 추정된 잡음 신호의 파워와 복소수 라플라시안 확률 통계 모델로부터 음성검출기(VAD)의 결정 규칙을 계산하는 가능성 비율 테스트(LRT) 계산부를 포함한다.

【대표도】

도 3

【색인어】

라플라시안 통계모듈, VAD, 가우시안

【명세서】**【발명의 명칭】**

복소수 라플라시안 통계모델을 이용한 음성 검출기 및 음성 검출 방법{A Voice Activity Detector Employing Complex Laplacian Model}

【도면의 간단한 설명】

도 1은 음성 스펙트럼의 라플라시안 누적 밀도 함수 및 가우시안 누적 밀도 함수를 경험적인 누적 밀도 함수와 비교한 곡선이다.

도 2는 라플라시안 모델과 가우시안 모델을 각각 적용한 음성 검출기의 수신기 동작 특성을 나타내는 도면이다.

도 3은 본 발명의 실시예에 따른 음성 검출기를 나타내는 도면이다.

【발명의 상세한 설명】**【발명의 목적】****【발명이 속하는 기술분야 및 그 분야의 종래기술】**

- <4> 본 발명은 음성 검출기 및 음성 검출 방법에 관한 것으로서, 특히 복소수 라플라시안 통계모델을 이용한 음성 검출기 및 음성 검출 방법에 관한 것이다.
- <5> 3GPP 및 3GPP2 등에서 제안되고 있는 많은 광대역 음성(speech) 코덱들이 가변율 전송(variable rate transmission) 기술을 요구한다. 이러한 가변율 전송을 위해, 음성 코덱(codec)은 음성이 없는 경우에 적은 비트를 할당할 수 있는 음성 검출기(voice activity detector)를 채택해야 한다. 즉, 음성 검출(voice Activity Detection, 이하 'VAD'라 함) 기술

이 가변율 음성 부호화(variable rate speech coding) 및 잡음 향상 기술에서 필수 불가결한 요소로 취급되고 있다.

<6> 최근 잡음이 섞인 음성(noisy speech; 이하 '잡음 음성'이라 함)으로부터 잡음과 음성을 분류하는 VAD 알고리즘의 성능을 향상시키기 위해 많은 방법들이 제안되고 있는데, 그 중 하나가 음성의 스펙트럼은 잡음의 스펙트럼보다 빨리 변한다는 직관(intuition)을 이용한 스펙트럼 비규칙성 측정 기반 모델(spectral irregularity measure-based model)이다. 그러나, 이 모델에 따르면 음성과 같은 스펙트럼을 가지는 잡음이 포함될 경우, 시스템의 성능이 극도로 떨어질 수 있다는 문제점이 있다.

<7> 한편, 통계적 모델을 채택하여 VAD의 성능을 향상시키기 위한 새로운 알고리즘이 'A statistical model-based voice activity detection', *IEEE Signal Processing Letters*, Vol. 6, No. 1 pp1-3, Jan. 1999 by J.Sohn, N.S. Kim and W.Sung의 논문(이하 '참조문헌 1'이라 함)에 제안되었다. 상기 모델은 가설의 집합에 적용되는 가능성 비율 테스트(likelihood ratio test; 이하 'LRT'라 함)로부터 VAD의 결정 규칙(decision rule)을 추출한다.

<8> VAD 알고리즘은 주로 이산 푸리에 변환(discrete Fourier transform; 이하 'DFT'라 함) 영역에서 동작하는데, 종래의 VAD 알고리즘은 복소수 가우시안 밀도(complex Gaussian density)에 의해 규정되는 순수 음성(clean speech)과 잡음의 스펙트럼 분포를 사용하였다.

<9> 그러나, DFT 계수를 복소수 가우시안 분포를 사용하여 순수 음성과 잡음을 모델링할 경우 정확성에 일정 정도 한계가 있었기 때문에, DFT 계수를 모델링할 새로운 분포 모델이 요구되고 있다.

【발명이 이루고자 하는 기술적 과제】

- <10> 본 발명은 복소수 라플라시안 모델을 채택한 음성 검출기 및 검출 방법을 제공하기 위한 것으로서, 라플라시안 모델과 가우시안 모델의 성능을 비교한다.

【발명의 구성 및 작용】

- <11> 상기한 목적을 달성하기 위한 본 발명의 특징에 따른 복소수 라플라시안 통계 모듈을 이용한 음성 검출기는
- <12> 입력된 음성에 대하여 고속 푸리에 변환을 수행하여, 시간영역의 음성신호를 주파수 영역에서 분석할 수 있도록 하는 고속 주파수 푸리에 변환기;
- <13> 상기 고속 주파수 푸리에 변환기로부터 출력되는 주파수 영역 상의 잡음에 오염된 음성($X(k)$)으로부터 잡음 신호의 파워($\lambda_{n,k}(t)$)를 추정하는 잡음 파워 추정부;
- <14> 상기 잡음 파워 추정부(20)로부터 추정된 잡음 신호의 파워 $\lambda_{n,k}(t)$ 와 복소수 라플라시안 확률 통계 모델로부터 VAD(voice activity detection)의 결정 규칙을 계산하는 가능성 비율 테스트(LRT) 계산부를 포함한다.
- <15> 또한, 본 발명의 특징에 따른 복소수 라플라시안 통계 모듈을 이용한 음성 검출 방법은
- <16> 잡음에 오염된 음성에 대하여 고속 푸리에 변환을 수행하여, 주파수 영역에서 분석할 수 있도록 출력하는 단계;
- <17> 상기 단계에서 출력되는 주파수 영역 상의 잡음에 오염된 음성($X(k)$)으로부터 잡음 신호의 파워($\lambda_{n,k}(t)$)를 추정하는 단계;
- <18> 상기 추정된 잡음 신호의 파워 $\lambda_{n,k}(t)$ 와 복소수 라플라시안 확률 통계 모델로부터 VAD의 결정 규칙을 계산하는 단계를 포함한다.

<19> 이하에서는 본 발명의 실시예를 도면을 참조하여 상세히 설명한다.

<20> 본 발명의 실시예는 다양한 잡음 조건에서 잡음에 오염된 음성신호의 DFT 계수를 VAD에 적용하기 위해 복소수 라플라시안 모델을 제안한다.

<21> 본 발명의 실시예에서는 먼저 다양한 잡음에 오염된 음성 조건하에서 적합 양호성 (Goodness Of Fit; 이하 'GOF'라 함) 테스트를 적용함으로써 라플라시안 모델과 가우시안 모델을 비교한 후, 가능성 비율 테스트(Likelihood Ration Test; 이하 'LRT'라 함)에 기반한 결정 규칙(Decision Rule)을 고려한다.

<22> 1. 통계적 모델

<23> 먼저, 잡음 신호 $X(t)$ 와 음성 신호 $S(t)$ 의 합이 $X(t)$ 라고 가정한다. 이때, 가설 H_0 와 H_1 은 각각 음성신호가 없는 경우와 음성 신호가 있는 경우를 나타낸다. 즉, 가설 H_0 와 H_1 에서, $X(t)$ 는 다음의 수학적 식 1과 수학적 식 2를 만족한다.

<24> **【수학적 식 1】** $H_0 : \text{speech absent} : X(t) = N(t)$

<25> **【수학적 식 2】** $H_1 : \text{speech present} : X(t) = N(t) + S(t)$

<26> 여기서, $X(t) = [X_0(t), X_1(t), \dots, X_{M-1}(t)]^T$, $N(t) = [N_0(t), N_1(t), \dots, N_{M-1}(t)]^T$, $S(t) = [S_0(t), S_1(t), \dots, S_{M-1}(t)]^T$ 는 각각 잡음 음성, 잡음 및 순수 음성의 이산 푸리에 변환(DFT) 계수이다.

<27> 상기한 통계적 모델은 적절한 DFT 계수 분포를 선택함으로써 완료된다. 본 발명의 실시예에서는 DFT 계수의 적절한 분포로서 가우시안 확률 밀도 함수(probabilistic density function, 이하 'PDF'라 함)를 채택하는 대신 복소수 라플라시안 확률밀도 함수를 채택하였다.

<28> 복소수 가우시안 PDF에서, 양 가설(H_0 , H_1)에 의해 결정되는 잡음 스펙트럼 요소(noisy spectral component)의 분포는 다음의 수학적 식 3과 수학적 식 4와 같이 주어진다.

<29> **【수학적 식 3】**
$$p_G(X_k|H_0) = \frac{1}{\pi\lambda_{n,k}} \exp\left\{-\frac{|X_k|^2}{\lambda_{n,k}}\right\}$$

<30> **【수학적 식 4】**
$$p_G(X_k|H_1) = \frac{1}{\pi[\lambda_{n,k} + \lambda_{s,k}]} \exp\left\{-\frac{|X_k|^2}{\lambda_{n,k} + \lambda_{s,k}}\right\}$$

<31> 여기서, $\lambda_{n,k}$ 와 $\lambda_{s,k}$ 는 각각 잡음 N_k 와 순수 음성 S_k 의 분산(variance)을 나타낸다.

<32> 한편, 복소수 라플라시안 PDF에서, DFT 계수 X_k 의 실수부(real part)와 허수부(imaginary part)를 각각 $X_{k(R)}$ 와 $X_{k(I)}$ 로 가정하면, $X_{k(R)}$ 와 $X_{k(I)}$ 는 다음의 수학적 식 5와 6과 같이 분포된다.

<33> **【수학적 식 5】**
$$p(X_{k(R)}) = \frac{1}{\sigma_x} \exp\left\{-\frac{2|X_{k(R)}|}{\sigma_x}\right\}$$

<34> **【수학적 식 6】**
$$p(X_{k(I)}) = \frac{1}{\sigma_x} \exp\left\{-\frac{2|X_{k(I)}|}{\sigma_x}\right\}$$

<35> 여기서, σ_x^2 는 X_k 의 분산을 나타낸다. X_k 의 실수부와 허수부가 독립적이라고 가정하면, 수학적 식 7과 같은 확률 밀도 함수 $p(X_k)$ 를 구할 수 있다.

<36> **【수학적 식 7】**
$$p(X_k) = p(X_{k(R)}) \cdot p(X_{k(I)}) = \frac{1}{\sigma_x^2} \exp\left\{-\frac{2(|X_{k(R)}| + |X_{k(I)}|)}{\sigma_x}\right\}$$

<37> 상기 수학적 식 7을 사용하여, 다음의 수학적 식 8과 9와 같은 잡음 DFT 계수의 분포를 구할 수 있다.

<38> **【수학적 식 8】**
$$p_L(X_k|H_0) = \frac{1}{\lambda_{n,k}} \exp\left\{-\frac{2(|X_{k(R)}| + |X_{k(I)}|)}{\sqrt{\lambda_{n,k}}}\right\}$$

<39>
$$p_L(X_k|H_1) = \frac{1}{\lambda_{n,k} + \lambda_{s,k}} \cdot \exp\left\{-\frac{2(|X_{k(R)}| + |X_{k(I)}|)}{\sqrt{\lambda_{n,k} + \lambda_{s,k}}}\right\}$$

【수학식 9】

<40> 성공적인 VAD 동작을 위해서, 본 발명의 실시예에서는 H_0 와 H_1 에 의해 결정되는 잡음 스펙트럼 요소를 위한 통계적 적합성 테스트(statistical fitting test)를 수행한다.

<41> 확률밀도함수(PDF)의 선택을 위해, 본 발명의 실시예에서는 GOF 테스트로서 잘 알려진 Kolomogorov-Sriminov 테스트(이하, 'KS 테스트'라 함)를 적용한다. KS 테스트를 적용함으로써, 각 통계적인 가정에 대한 신뢰성 있는 관측을 보장할 수 있다.

<42> KS 테스트는 경험적인(empirical) 누적 분포 함수(cumulative distribution function; 이하 'CDF'라 함) F_X 를 소정의 분포 함수 F 와 비교한다. 경험적 CDF는 R.C. Reininger와 D.Gibson이 발표한 논문(이하, '참조논문 2'이라 함)인 'Distributions of the two dimensional DCT coefficients for images,' *IEEE Trans. Communications.*, Vol. Com-31, No. 6, June 1983를 사용하였다.

<43> 상기 논문에 기초한 경험적 CDF는 $\mathbf{X} = [X_0, X_1, \dots, X_{N-1}]^T$ 가 잡음음성의 DFT 계수를 나타내는 벡터라고 가정할 때, 다음의 수학식 10으로 나타낼 수 있다.

<44>
$$F_X(z) = \begin{cases} 0, & z < X_{(1)} \\ \frac{n}{N}, & X_{(n)} \leq z < X_{(n+1)}, \quad n = 0, 1, \dots, N-1 \\ 1, & z \geq X_{(N)} \end{cases}$$

【수학식 10】

<45> 여기서, $X_{(n)}, n = 0, \dots, N-1$ 은 데이터 X 의 순서 통계(order statistic)이다. 이러한 순서 통계를 계산하기 위해, 본 발명의 실시예에서는 X 의 원소(element)를 분류하여, X 의 가장 작은 원소인 $X_{(0)}$ 에서부터 가장 큰 원소인 $X_{(N-1)}$ 까지 순서대로 배열한다.

- <46> 잡음 환경을 모의 실험하기 위해, 서로 다른 4명의 남성과 여성 화자로부터, 64초구간의 음성 자료(speech material)를 수집하였으며, NOISEX-92 데이터베이스로부터 추출한 백색 잡음(white noise)과 차량 잡음(vehicular noise)을 신호대 잡음비(SNR)가 10dB인 순수 음성 신호에 첨가하였다. 수집한 상기 데이터를 사용하여, 표본 평균과 분산을 계산하였으며, 주어진 라플라시안 분포와 가우시안 분포에 적용하였다.
- <47> 도 1은 잡음 음성 스펙트럼(실수부)의 라플라시안 CDF 및 가우시안 CDF를 경험적인 CDF와 비교한 그래프로서, (a)와 (b)는 각각 H_1 이 백색 잡음(SNR=10dB)과 차량 잡음(SNR=20)인 경우를 나타낸다.
- <48> 도 1로부터, 백색 잡음 및 차량 잡음 환경 모두에서 라플라시안 CDF의 곡선이 가우시안 CDF의 곡선 보다 경험적인 CDF 곡선에 가깝다는 것을 알 수 있다.
- <49> 경험적인 CDF와 주어진 분포 사이의 거리 측정을 규정하기 위해, 본 발명의 실시예에서는 상기 참조논문 2에서 제안한 KS 테스트 통계량(statistic)을 사용한다.
- <50> KS 테스트 통계량(T)는 다음의 수학적 식 11과 같이 정의된다.
- <51> **【수학적 식 11】**
$$T = \max_i |F_X(X_i) - F(X_i)|$$
- <52> 여기서, 표본 포인트 $\{X_i\}$ 에서 구한 $F_X(X_i)$ 와 $F(X_i)$ 사이의 최대 차이가 거리에 해당한다.
- <53> 몇몇 분포에 대하여 데이터를 테스트하는 경우, 최소의 KS 통계량을 나타내는 분포는 주어진 데이터에 가장 적합한 것으로 고려될 수 있다. 표 1은 다양한 잡음 환경하의 잡음 음성의 DFT 계수를 위한 KS 테스트의 결과를 나타낸다. 표 1에서 G와 L은 각각 가우시안 분포와 라플라시안 분포를 나타낸다.

<54> 【표 1】

noise		white			vehicular			babble		
SNR(dB)		5	10	15	5	10	15	5	10	15
H_1	$G; X_{k(R)}$	0.043	0.078	0.129	0.211	0.223	0.231	0.129	0.165	0.198
	$L; X_{k(R)}$	0.031	0.025	0.068	0.164	0.177	0.186	0.071	0.107	0.145
	$G; X_{k(I)}$	0.044	0.081	0.134	0.214	0.225	0.232	0.142	0.173	0.203
	$L; X_{k(I)}$	0.028	0.026	0.073	0.164	0.178	0.187	0.080	0.116	0.149
H_0	$G; X_{k(R)}$	0.045	0.052	0.063	0.238	0.270	0.311	0.149	0.127	0.136
	$L; X_{k(R)}$	0.024	0.024	0.023	0.189	0.237	0.277	0.088	0.167	0.078
	$G; X_{k(I)}$	0.051	0.059	0.071	0.243	0.275	0.325	0.153	0.127	0.134
	$L; X_{k(I)}$	0.019	0.016	0.021	0.243	0.237	0.278	0.093	0.067	0.075

<55> 표 1로부터, 모든 잡음 환경 하에서 라플라시안 모델의 KS 통계량(T)이 가우시안 모델의 KS 통계량 보다 적음을 알 수 있다. 따라서, 라플라시안 모델이 가우시안 모델보다 DFT 계수를 모델링하는데 훨씬 정확하다는 것을 알 수 있다.

<56> 2. LRT에 기초한 결정 규칙

<57> 본 발명의 실시예는 가정된 통계적 모델에 기초하여, k번째 주파수 빈(bin)에 대한 가능성 비율(likelihood ratio; 이하 'LR'이라 함)을 수학적 12와 같이 계산하였다.

<58>
$$\Lambda_k \equiv \frac{p(X_k|H_1)}{p(X_k|H_0)}.$$
 【수학적 12】

<59> VAD를 위한 결정 규칙(decision rule)은 개별 주파수 채널에 대한 LR의 기하학적 평균으로 정의될 수 있으며, 이것은 다음의 수학적 13과 같이 나타낼 수 있다.

<60>
$$\log \Lambda = \frac{1}{M} \sum_{k=0}^{M-1} \log \Lambda_k \underset{H_0}{\overset{H_1}{>}} \eta$$
 【수학적 13】

<61> 여기서, η 은 결정을 위한 문턱값이다.

<62> H_0 및 H_1 에 대한 종래의 가우시안 분포에서 LR은 다음의 수학적 14에 의해 구해진다.

<63>
$$\Lambda_k^{(G)} = \frac{p_G(X_k|H_1)}{p_G(X_k|H_0)} = \frac{1}{1 + \xi_k} \exp \left\{ \frac{\gamma_k \xi_k}{1 + \xi_k} \right\}$$

【수학식 14】

<64> 여기서, $\xi_k = \lambda_{s,k}/\lambda_{n,k}$ 이고, $\gamma_k = |X_k|^2/\lambda_n$ 이다.

<65> 한편, 라플라시안 모델에 기초하여 계산된 LR은 다음의 수학식 15와 같다.

<66>
$$\Lambda_k^{(L)} = \frac{p_L(X_k|H_1)}{p_L(X_k|H_0)} = \frac{1}{1 + \xi_k} \exp \left\{ 2(|X_{k(R)}| + |X_{k(I)}|) \left(\frac{|X_k| - \sqrt{\lambda_{n,k}}}{\sqrt{|X_k| \lambda_{n,k}}} \right) \right\}$$

【수학식 15】

<67> 여기서, VAD의 성공 또는 실패 여부는 통계적인 모델뿐만 아니라 잡음 파워(noise power) $\{\lambda_{n,k}(t)\}$ 와 음성 파워(speech power) $\{\lambda_{s,k}(t)\}$ 에 대한 적절한 추정에 의해서도 결정된다.

<68> 3. 실험 결과

<69> 본 발명의 실시예에서는 라플라시안 모델의 성능과 가우시안 모델의 성능을 비교하기 위해, 각 통계적 모델에 대한 음성 감지 확률(P_d)과 오류(false alarm) 확률(P_f)을 조사하였다.

<70> 도 2는 5dB SNR에서 라플라시안 모델과 가우시안 모델을 적용한 VAD의 수신기 동작 특성을 나타내는 그래프로서, (a)와 (b)는 백색 잡음과 차량 잡음인 경우를 나타낸다. 도 2에서, 세로축과 가로축이 각각 음성 감지 확률(P_d)과 오류(false alarm) 확률(P_f)을 나타낸다.

<71> 도2에 도시한 수신기 동작 특성으로부터, 두 통계 모델의 P_d 와 P_f 사이에는 트레이드-오프가 존재한다는 것을 알 수 있으며, 음성 감지 확률(P_d)이 정상 영역(90% 이상)에 있으면, 복소수 라플라시안 모델 기반의 결정 규칙인 복소수 가우시안 기반의 결정 규칙보다 더 좋다는 것을 알 수 있다.

<72> 이상에서 설명한 바와 같이, 다양한 잡음 환경에서 복소수 라플라시안 모델 기반의 VAD가 복소수 가우시안 기반의 VAD 보다 더 좋은 성능을 나타낸다는 것을 알 수 있다.

- <73> 다음은 본 발명의 실시예에 따른 복소수 라플라시안 모델을 적용한 음성 검출기에 대하여 설명한다.
- <74> 도 3은 본 발명의 실시예에 따른 음성 검출기를 나타내는 도면이다.
- <75> 도 3에 도시한 바와 같이, 본 발명의 실시예에 따른 음성 검출기는 고속 푸리에 변환기(10, Fast Fourier Transformer; FFT), 잡음 파워 추정부(20), LRT 계산부(30)를 포함한다.
- <76> FFT(10)는 입력된 음성을 고속 푸리에 변환하여, 음성신호를 주파수 영역에서 분석할 수 있도록 잡음에 오염된 음성($X(k)$)을 출력한다. 잡음 파워 추정부(20)는 상기 FFT(10)로부터 출력되는 주파수 영역 상의 잡음에 오염된 음성($X(k)$)으로부터 잡음 신호의 파워를 추정한다. LRT(Likelihood Ratio Test) 계산부(30)는 상기 잡음 파워 추정부(20)로부터 추정된 잡음 신호의 파워 $\lambda_{n,k}(t)$ 와 기정의된 음성신호의 존재가설(H_0 , H_1)에 대한 복소수 라플라시안 확률 통계 모델로부터 VAD의 결정 규칙을 계산한다.
- <77> 이때, 결정 규칙은 앞서 설명한 바와 같이, 개별 주파수 빈에 대한 가능성 비율들(Likelihood ratios)에 대한 기하 평균으로 정의되며, 라플라시안 모델의 가능성 비율은 수식 15에 의해 결정된다.
- <78> 이상에서 본 발명의 실시예에 대하여 설명하였으나, 본 발명은 상기한 실시예에만 한정되는 것은 아니며, 그 외의 다양한 변경이나 변형이 가능하다.

【발명의 효과】

- <79> 이상에서 설명한 바와 같이, 본 발명의 음성 검출기는 라플라시안 통계 분포를 사용하기 때문에, 복소수 가우시안 기반의 VAD 보다 더 좋은 성능을 나타낼 수 있다.

【특허청구범위】

【청구항 1】

입력된 음성에 대하여 고속 푸리에 변환을 수행하여, 시간영역의 음성신호를 주파수 영역에서 분석할 수 있도록 하는 고속 주파수 푸리에 변환기;

상기 고속 주파수 푸리에 변환기로부터 출력되는 주파수 영역 상의 잡음에 오염된 음성($X(k)$)으로부터 잡음 신호의 파워($\lambda_{n,k}(t)$)를 추정하는 잡음 파워 추정부;

상기 잡음 파워 추정부(20)로부터 추정된 잡음 신호의 파워 $\lambda_{n,k}(t)$ 와 복소수 라플라시안 확률 통계 모델로부터 VAD(voice activity detection)의 결정 규칙을 계산하는 가능성 비율 테스트(LRT) 계산부를 포함하는 복소수 라플라시안 통계 모듈을 이용한 음성 검출기.

【청구항 2】

제1항에 있어서,

상기 결정 규칙은 k 번째 주파수에 대한 가능성 비율(Λ_k)의 기하학적 평균이며, 상기 가능성 비율 Λ_k 은 다음의 식

$$\Lambda_k \equiv \frac{p(X_k|H_1)}{p(X_k|H_0)},$$
 - 가설 H_0 와 H_1 은 각각 음성신호가 없는 경우와 음성 신호가 있는 경우를 나타내며, X_k 는 k 번째 이산 푸리에 계수를 나타냄-으로 결정되는 것을 특징으로 하는 복소수 라플라시안 통계 모듈을 이용한 음성 검출기.

【청구항 3】

제2항에 있어서,

상기 라플라시안 통계 모듈을 적용한 가능성 비율은 다음의 수학식으로

$$\Lambda_k^{(L)} = \frac{p_L(X_k|H_1)}{p_L(X_k|H_0)} = \frac{1}{1 + \xi_k} \exp \left\{ 2(|X_{k(R)}| + |X_{k(I)}|) \left(\frac{|X_k| - \sqrt{\lambda_{n,k}}}{\sqrt{|X_k| \lambda_{n,k}}} \right) \right\} \text{여기서, } \xi_k = \lambda_{s,k} / \lambda_{n,k} \text{이고,}$$

$X_{k(R)}$ 와 $X_{k(I)}$ 는 X_k 의 실수부와 허수부를 나타냄-결정되는 것을 특징으로 하는 복소수 라플라시안 통계 모듈을 이용한 음성 검출기.

【청구항 4】

입력된 음성에 대하여 고속 푸리에 변환을 수행하여, 시간영역의 음성신호를 주파수 영역에서 분석할 수 있도록 잡음 음성($X(k)$)을 출력하는 단계;

상기 단계에서 출력되는 주파수 영역 상의 잡음 음성($X(k)$)으로부터 잡음 신호의 파워($\lambda_{n,k}(t)$)를 추정하는 단계;

상기 추정된 잡음에 오염된 신호의 파워 $\lambda_{n,k}(t)$ 와 복소수 라플라시안 확률 통계 모델로부터 VAD의 결정 규칙을 계산하는 단계를 포함하는 복소수 라플라시안 통계 모듈을 이용한 음성 검출방법.

【청구항 5】

제4항에 있어서,

상기 결정 규칙은 k 번째 주파수에 대한 가능성 비율의 기하학적 평균이며, 상기 가능성 비율은 다음의 식

$$\Lambda_k^{(L)} = \frac{p_L(X_k|H_1)}{p_L(X_k|H_0)} = \frac{1}{1 + \xi_k} \exp \left\{ 2(|X_{k(R)}| + |X_{k(I)}|) \left(\frac{|X_k| - \sqrt{\lambda_{n,k}}}{\sqrt{|X_k| \lambda_{n,k}}} \right) \right\} \text{여기서, 가설 } H_0 \text{와 } H_1 \text{은 각각}$$

음성신호가 없는 경우와 음성 신호가 있는 경우를 나타내며, X_k 는 k 번째 이산 푸리에 계수, $\xi_k = \lambda_{s,k} / \lambda_{n,k}$, $X_{k(R)}$ 와 $X_{k(I)}$ 는 X_k 의 실수부와 허수부를 나타냄-으로 결정되는 것을 특징으로 하는 복소수 라플라시안 통계 모듈을 이용한 음성 검출 방법.

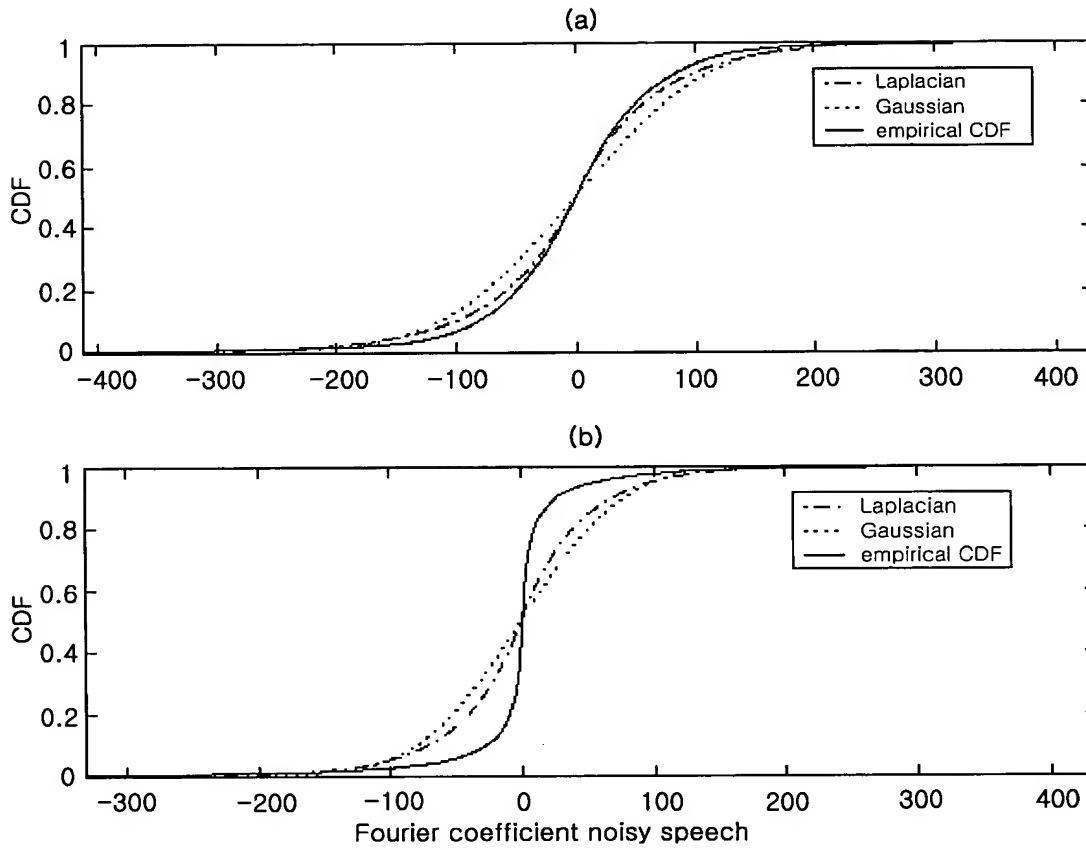


1020020083728

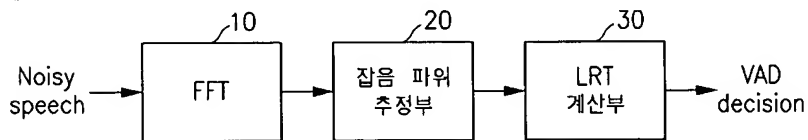
출력 일자: 2003/7/28

【도면】

【도 1】



【도 3】



【도 2】

